基于深度学习的单图像超分辨率重建研究综述 *

南方哲，钱育蓉，行艳妮，赵京霞
(新疆大学 软件学院，乌鲁木齐 830046)

摘　要：为深入了解基于深度学习的单图像超分辨率重建(single image super-resolution，SISR)的发展，把握当前研究的热点和方向，针对现有基于深度学习的单图像超分辨率重建模型进行了梳理。首先介绍了相关深度学习算法；然后介绍了基于深度学习的模型以及评价指标，并通过实验对比分析现有模型的性能，其目的在于从本质上了解基于深度学习的单图像超分辨率重建模型的优势；最后对单图像超分辨率重建的关键问题进行了总结，并对未来的发展趋势进行了展望。

关键词：单图像超分辨率重建；深度学习；密集卷积网络；生成式对抗网络
中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2018.10.0610

Survey of single image super resolution based on deep learning

Nan Fangzhe, Qian Yurong, Xing Yanni, Zhao Jingxia
(*College of Software, Xinjiang University, Urumqi* 830046, *China*)

Abstract: In order to understand the development of single image super-resolution reconstruction (SISR) based on deep learning and grasp the hotspots and directions of the current research, this paper combs the existing model of single image super-resolution reconstruction based on deep learning. Firstly, the paper introduces the related deep learning algorithm, these models based on deep learning and their evaluation index. In addition, it compares the performance of existing models through experiments, which aims to understand the advantages of single-image super-resolution reconstruction model based on deep learning. Finally, the paper summarizes the key issues of single-image super-resolution reconstruction, and prospects the future development trends.

Key words: single image super-resolution; deep learning; densenet; generative adversarial networks

0　引言

图像超分辨率重建就是指由一张低分辨率(low-resolution，LR)图像或者图像序列恢复出高分辨率(high-resolution, HR)图像的技术[1]。高分辨率图像像素密度高，可获取细节多，在实际应用中不可或缺。目前超分辨率重建(Super-Resolution, SR)可分为两类，分别是从多张低分辨率图像重建出高分辨率图像和从单张低分辨率图像重建出高分辨率图像。由单一 LR 图像生成 HR 图像的技术已被广泛应用于高光谱成像、医学成像、卫星遥感等方面，其应用的灵活简便及实用性，使得单图像超分辨率重建得到了广泛的关注。本文将主要介绍基于单张低分辨率的重建方法(single image super-resolution, SISR)。

传统的 SISR 的方法包括基于插值的算法，如双三插值法和 Lanczos 重采样算法[2]；基于重建的算法，如迭代反投影 IBP[3]、最大后验概率(maximum a posterior, MAP)[4]、凸集投影法(projection onto convex sets，POCS)[5]；以及基于学习的算法，如邻居嵌入和局部线性嵌入[6]方法、稀疏编码[7]和稀疏编码网络[8]。这些方法或利用相同图像的内部相似，或学习外部低分辨率和高分辨率样本对的映射函数，虽然注重学习和优化字典，但是上述方法的其余步骤很少在统一优化框架中进行优化或考虑。

由于深度学习在其他计算机视觉领域中取得的突破性进展，人们尝试引入深度神经网络，通过构建深层次的网络进行端到端的训练来解决图像超分辨率重建问题。如文献[9]将多层协作自动编码器堆叠在一起，以实现自相似贴片的稳健匹配。深度卷积神经网络[10]和反卷积网络[11]被设计为以类似于耦合稀疏编码的方式直接学习从 LR 空间到 HR 空间的非线性映射[12]。

1　深度学习

1.1　深度学习概念

深度学习(deep learning，DL)源于人工神经网络，其概念由 Hinton 等人[13]在 2006 年提出，由此开启了深度学习在学术界和工业界的浪潮。DL 作为机器学习研究的新领域，通过模仿人脑的机制来解释数据，让机器自动学习良好的特征，免去人工选取过程[14]。目前深度学习在模型、算法、硬件设施与开发社区四方面已经取得了重大突破[15]，解决了以往神经网络优化困难、应用受限、计算缓慢、认可度不高等问题。目前在单图像超分辨率重建问题中常用的深度学习网络有卷积神经网络、深度残差网络、循环神经网络、密集卷积网络、生成式对抗网络等。

收稿日期：2018-10-31；**修回日期：**2018-11-23　**基金项目：**国家自然科学基金资助项目（61562086，61462079）；新疆维吾尔自治区教育厅创新团队项目(XJEDU2016S035)

作者简介：南方哲（1994-），女，河南周口人，硕士研究生，主要研究方向为计算机视觉、单图像超分辨率重建等（1508632217@qq.com）；钱育蓉（1980-），女，教授，博士，主要研究方向为网络计算、遥感图像处理；行艳妮（1994-），女，硕士研究生，主要研究方向为数据挖掘、算法优化等；赵京霞（1995-），女，硕士研究生，主要研究方向为图像处理、模式识别．

1.2 深度学习模型介绍

1.2.1 卷积神经网络

近年来，图像特征的提取与选择一直是计算机视觉领域的一个基础而重要的研究方向，卷积神经网络(convolutional neural network，CNN)提供的端到端的学习模型可以通过传统的梯度下降方法训练学习图像特征[16]，相比于其他神经网络方法，CNN 更适合应用于图像特征的学习与表达。目前 CNN 已成为计算机视觉中深度学习算法的关键。CNN 最早可追溯到 15 年前[17],近年来由于它在图像分类上的成功[18]，出现了爆炸性的流行。CNN 主要由输入层、卷积层、池化层、全连接层组成，其结构如图 1 所示。再加之功能强大的 GPU 的有效应用[19]、激活函数 ReLU[20]的出现，使得 CNN 在保持良好训练效果的同时收敛速度更快，可以轻松地访问大量数据(如 ImageNet [21])，因此基于 CNN 的单图像超分辨率重建方法也在不断进步。

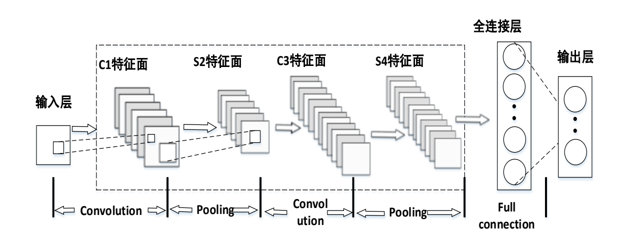


图 1 卷积神经网络的概念结构图

Fig.1 Conceptual structure diagram of CNN

1.2.2 深度残差网络

理论上在特征提取时增加网络的深度或者宽度，提取到的细节与抽象能力就越丰富，然而仅仅进行简单的层数增加操作，会出现梯度爆炸或梯度弥散的问题。传统解决方法如数据初始化(normlized initializatiton)和正则化（batch normlization）操作，虽然解决了梯度问题，但是会导致网络性能退化。残差(ResNet)网络[22]则克服了神经网络深度增加导致性能易退化的缺点，其主要思想是在标准的前馈卷积网络上，添加一个跳跃得以绕过一些层的连接，使网络模型能够在保证网络性能的同时，通过增加深度来提高网络性能。其结构如图 2 所示。

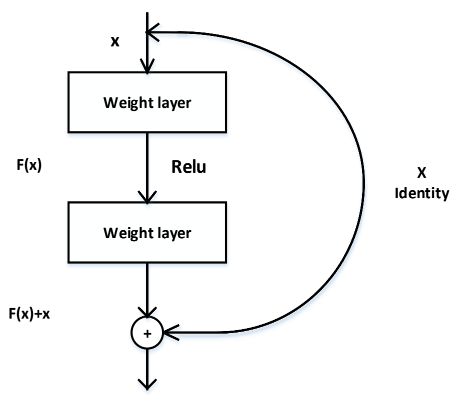


图 2 残差网络结构

Fig.2 Structure diagram of residual network

在 SISR 问题中输入的低分辨率图像和输出的高分辨率图像在很大程度上是相似的，即低分辨率图像携带的低频信息与高分辨率图像的低频信息相近。为节约训练时间，只需要学习高分辨率图像与低分辨率图像之间的高频部分残差即可。残差网络结构的思想特别适合用来解决此问题，且残差网络已经被广泛应用到计算机视觉[23]、图像识别[24]、行人监测[25]等相关领域的模型中。

1.2.3 递归(循环)神经网络

循环神经网络 (recurrent neural networks，RNN) [26]中神经元的输出可以在下一个时间戳直接作用到自身，克服了全连接的 CNN 网络中存在的问题，如每层神经元的信号只能向上一层传播、样本的处理在各个时刻相互独立、无法对时间序列进行建模等。RNN 的基本思想是：将上一个时刻的输出和当前时刻的输入同时作为当前网络输入，从而得到该时刻的输出，然后不断迭代上述过程。对比 CNN 通过卷积运算共享权重从而减少计算量的思想，RNN 从头到尾所有的权重都是公用的，不同的只是输入和上一时刻的输出。RNN 已经成功运用到机器翻译[27]、语音识别[28]、图像自动生成[29]等领域。研究者们经实验发现 RNN 在加深网络深度的同时，也可提高网络的准确率。RNN 结构如图 3 所示。

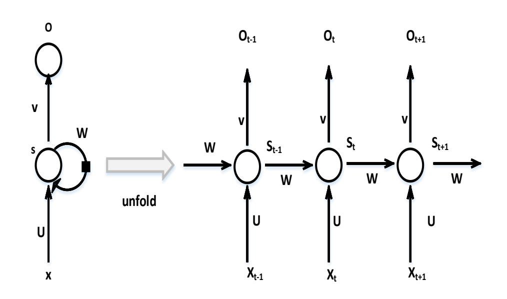


图 3 RNN 结构

Fig.3 Structure diagram of RNN

1.2.4 生成式对抗网络

生成式对抗网络(generative adversarial networks，GAN)是 Goodfellow 等人[30]在 2014 年提出的一种生成式模型。结构如图 4 所示。其思想受博弈论中的二人零和博弈(即二人的利益之和为零，一方的所得正是另一方的所失)的启发，由一个生成器和一个判别器构成。生成器捕捉真实数据样本的潜在分布，并生成新的数据样本；判别器是一个二分类器，判别输入的是真实数据还是生成数据。生成器和判别器均可采用深度神经网络模型实现[31]。GAN 发展至今，其家族在不断扩大，常用的模型有 DCGAN[32]、InfoGAN[33]、EBGAN[34]、Improved GAN[35]、WGAN[36]。目前 GAN 已经在图像处理[37]，计算机视觉[38]、语音识别[39]等方面得到了应用。相较于其他方法，加入 GAN 算法的模型能充分利用图像信息，所以模型的准确率均有提高。

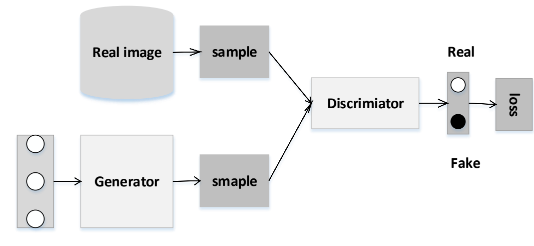


图 4 GAN 结构

Fig. 4 Structure diagram of GAN

1.2.5 密集卷积网络

在深度学习网络中，随着网络深度的增加，梯度消失问题会愈加明显。目前很多论文都针对这个问题提出了解决方案，如 ResNet、Highway Networks[40]、Stochastic depth[41]、FractalNets[42]等。尽管这些算法的网络结构有差别，但是核心都在于创建从早期图层到后期图层的短路径。考虑到在保证网络中层与层之间最大程度的信息传输的前提下，直接将所有层连接起来则可以进行有效训练， Huang 等人[43]提出了密集卷积网络(dense convolutional network，DenseNet)。该

网络模型的主要思想为：假设模型有 L 层，传统的神经网络就会有 L 个连接，而在 DenseNet 中，会有 $L(L+1)/2$ 个连接，即每层的输入都来自前面所有层的输出。网络结构如图 5 所示。该网络可以缓解消失梯度问题，加强特征传播，鼓励特征重用，并大幅减少参数数量。DenseNet 一经提出得到了广泛的应用，如语义切割[44]、语音识别[45]和图像分类[46]等。

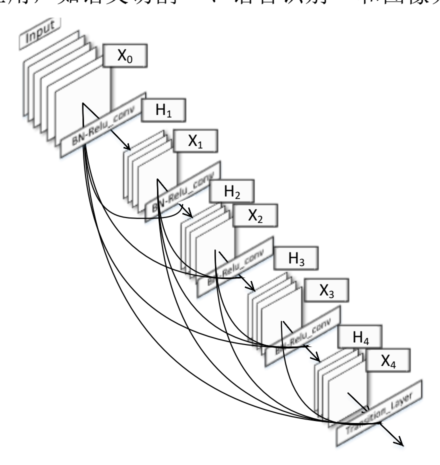


图 5　DenseNet 结构

Fig.5　Structure diagram of densenet

2　基于深度学习的单图像超分辨率模型

目前常用的深度学习的模型可根据图像的输入信息、训练过程和特征提取以及高频细节的应用等将基于深度学习的单图像超分辨率模型分为基于插值预处理的模型、基于原始图像处理的模型、基于分层特征的模型和基于高频细节的模型四大类。

2.1　基于插值预处理的模型

基于插值预处理的模型的主要思想为：原始图像输入前需经过插值预处理把图像放大到规定目标尺寸。现有的基于插值预处理的模型所用到的深度学习网络有 CNN、RNN 和残差网络以及三者的集合。

SRCNN(super-resolution convolutional neural network)[47]是最早运用深度学习方法在 LR 与 HR 之间建立端到端映射的 SISR 模型，其输入图像采用了插值预处理的方法。SRCNN 的结构简单，仅用了三个卷积层，分别对应图像块的提取和特征表示、特征非线性映射和重建。由于均方误差是数据序列与真实值之间的关系，采用均方误差作为该模型的损失函数，如式（1）所示。

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\|F(Y_i;\theta) - X_i\right\|^2 \tag{1}$$

SRCNN 较传统的模型(SC、K-SVD、ARN)在相同的数据结构上放大相同倍数其峰值信噪比和运算速度都有提升。

VDSR(accurate image super-resolution using very deep convolutional networks)是由 Kim 等人[48]提出的一种基于非常深的卷积网络的高度准确的方法，该方法将插值预处理的图像作为输入，通过将卷积层与残差网络相叠来增加网络深度(20 层)，使得越深的网络层拥有更大的感受野；同时利用残差学习和自适应梯度裁剪(adjustable gradient clipping)加快收敛过程，在卷积前填充零，以保持所有要素图（包括输出图像）的大小相同，模型解决了不同倍数的超分辨率问题。由于 VDSR 模型的输入和输出相似，其损失函数如式（2）所示。

$$\frac{1}{2}\left\|r - f(x)\right\|^2 \tag{2}$$

基于插值预处理的方法还有 DRCN[49]、DRRN[50]、IRCNN[51]、Memnet[52]和 SDSR[53]等。其中 DRCN 和 DRRN 是在原有 CNN 模型的基础上加入了 RNN 的思想，增加网络的深度，提高了模型的效率。

2.2　基于原始图像处理的模

基于插值的深度学习方法最大缺点是原始 LR 图像需要经过插值预处理，这个预处理步骤不仅容易平方地增加计算复杂度，而且会导致原始 LR 图像过度平滑和模糊，更易丢失一些细节。目前研究者们主要在现有基于插值预处理的模型的基础上利用反卷积层和子像素卷积层对模型结构进行改进，直接对原始图像进行处理来解决插值预处理问题。

反卷积网络可以看做是卷积层的逆过程，假设步长为 n，可以把原始图像尺寸放大 n 倍，实现上采样的操作。反卷积层用于解决插值预处理过程，如 Dong 等人[54]针对图像超分辨重建插值预处理问题基于 CNN，利用反卷积层的方法提出了 FSRCNN(fast super-resolution convolutional neural networks)模型。FSRCNN 直接将原始的低分辨率图像输入到网络中，使用更小的卷积核和更多的映射层，且映射层是可共享的，可以训练不同上采样倍率。FSRCNN 可以分为特征提取、缩小、映射、扩展和反卷积五个部分。前四个部分是卷积层，而最后一个是反卷积层。该模型与 SRCNN 一样采用 MSE 作为损失函数，优化的目标表示如式（3）所示。

$$\min_{\theta}\frac{1}{n}\sum_{i=1}^{n}\left\|F(Y_s^i;\theta) - X^i\right\|_2^2 \tag{3}$$

其中：X^i、Y^i 表示训练样本中第 i 个 LR 和 HR 对应的图。RED[55]采用对称的卷积层—反卷积层网络结构，作为一个编码—解码框架，直接学习由低分辨率图像到高分辨率图像端到端的映射。LapSRN[56]采用卷积层和反卷积层交替连接的方式，构造一个多级的网络解决基于插值的问题。

子像素卷积层是另一种解决插值预处理问题的方法。其思想将最后一层卷积层输出的特征个数设置成固定值，即放大倍数 r 的平方，保证总的像素个数与要得到的高分辨率图像一致，将像素进行重新排列得到高分辨率图。例如 Wenzhe Shi 等人[57]提出一种新颖的 CNN 架构 ESPCN，其在 LR 空间中直接提取特征图，通过引入有效的子像素卷积层，学习升级滤波器阵列，将最终的 LR 特征图放大到 HR 中。该方法有效地将 SR 管道中的手工双立方滤波器替换为针对每个特征映射专门训练的更复杂的放大滤波器，同时降低了整个 SR 操作的计算复杂度。ESPCN 的损失函数为

$$l(W_{1:L}, b_{1:L}) = \frac{1}{r^2HW}\sum_{x=1}^{rH}\sum_{x=1}^{rW}\left(I_{x,y}^{HR} - f_{x,y}^{L}(I^{LR})\right)^2 \tag{4}$$

SRResNet[58]和 EDSR [59]同样采用了有效的亚像素卷积层，并利用了残差学习在 LR 空间中提取特征，使用子像素卷积层对最终 LR 特征进行放大。

2.3　基于分层特征学习的模型

大多数深度 SISR 模型没有充分利用原始低分辨率(LR)图像的分层特征，导致性能相对较低。密集网络(DenseNet)因其基于分层特征学习的优点，被应用到 SISR 中解决此问题。

例如 SRDenseNet[60]充分运用了 DenseNet 的思想，首先是用一个卷积层学习低层的特征，接着用多个稠密块学习高层的特征，然后通过几个反卷积层学到上采样滤波器参数，最后通过一个卷积层生成高分辨率输出。该模型的损失函数如式（5）所示。

$$l(\theta) = \frac{1}{N}\sum_{K=1}^{N}\left\|F(I_L^K,\theta) - I_H^K\right\|_2^2 \tag{5}$$

该模型在 SISR 问题上取得了不错的结果。

Zhang 等人[61]提出了一种新的残差密集网络(residual dense network，RDN)，该网络的灵感同样来自于 DenseNet，该模型充分利用原始 LR 图像中的所有分层特征。其主要思想为：利用残差密集块(RDB)来提取密集连接卷积层的丰富局部特征；然后使用 RDB 中的局部特征融合自适应地学习来自先前和当前局部特征的更有效特征，并提高在更广泛网络训练时的稳定性；在完全获得密集的局部特征之后，使用全局特征融合，以整体方式联合和自适应地学习全局分层特征；通过全局残差学习，将浅特征和深特征结合在一起，从而产生原始 LR 图像的全局密集特征。

Zhou 等人[62]基于 DenseNet 结合高频信息提出了利用高频信息增强密集连接卷积神经网络(SRDN)的超分辨率模型，使网络更加注重训练过程中边缘和纹理等高频区域的重建。

2.4 基于高频细节特征的模型

尽管现有模型使用更快更深的神经网络在 SISR 的准确性和速度方面取得了突破，产生的估计值具有较高的峰值信噪比，但当出现大规模的放大时，通常缺乏高频细节，并且在无法与较高分辨率预期的保真度相匹配，因此解读图像的细节是一个亟待解决的问题。

生成式对抗网络常用于解决此类问题。例如 Ledig 等人[63]基于 GAN 提出了一种用于图像超分辨率的生成对抗网络 SRGAN，通过生成式和判别器的交替执行，充分提取高频信息。它是第一个能够推断在放大 4 倍的情况下，照片依然逼真的框架。该模型的主要思想是提出了一种感知损失函数，公式为

$$\underbrace{I^{SR} = \underbrace{I_X^{SR}}_{content\ \ loss} + \underbrace{10^{-3} I_{Gen}^{SR}}_{adversarial\ \ loss}}_{perceptual\ \ loss(for\ \ VGG\ \ based\ \ content\ \ loss)} \tag{6}$$

它包括内容丢失和对抗性损失。其中内容损失可以是基于均方误差的损失函数：

$$l_{MSE}^{SR} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}\left(I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y}\right)^2 \tag{7}$$

也可以是基于训练好的以 ReLU 为激活函数的 VGG 模型的损失函数:

$$l_{VGG_{i,j}}^{SR} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\left(\varphi_{i,j}(I^{HR})_{x,y} - \varphi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y}\right)^2 \tag{8}$$

对抗性损失则使用鉴别器网络将解决方案推向自然图像流形,该网络被训练以区分超分辨图像和原始照片真实图像。对抗损失为

$$l_{Gen}^{SR} = \sum_{n=1}^{N} = \log_{D_{\theta_D}}(G_{\theta_G}(I^{LR})) \tag{9}$$

另外，该模型使用由感知相似性驱动的语义丢失来替换像素空间中的相似性方法,视觉效果较其他方法有显著提高。文献[64]将 DenseNet 作为 GAN 网络的体系结构解决了图像超分辨率问题。Zhang 等人[65]利用 GAN 从 LR 中生成了清晰而真实的 HR 图像。

利用在线检索的信息是高频信息补偿的另一种方法。例如杨文瀚等人[66]为了弥补歧义性造成的高频细节损失，提出了一种基于深度网络、利用在线检索的数据进行高频信息补偿的图像超分辨率重建算法。该方法通过三条分支预测高分辨率重建结果来构建一个深度网络：一条旁路直接将输入的低分辨率图像输入到网络的最后一层；一条内部高频信息重建路径基于低分辨率图像回归预测高分辨率图像，重建高分辨率图像的主要结构；另一条外部高频信息补偿路径根据内部重建的结果，从在线检索到的相似图像中提取高频细节，对内部重建的结果进行细节补偿。

3 实验结果与分析

本节将从实验的角度对几种基于深度学习的单图像超分辨率模型进行比较分析，首先给出了模型的性能评价指标，然后在不同数据库上比较和分析几种单图像超分辨率重建模型的性能。

3.1 图像质量评估标准

SISR 模型的评价通常分为主观评价和客观评价两大类。主观评价即征求人的意见，由人决定图像分辨率的高低。主观评价方法目前的评价标准常用的方法为平均主观意见分 MOS (mean opinion score) [69]，即假设打分规则为“最差”到“优秀”五个等级。参与者根据所看到的的图像与原图像的差别给出等级数。

为了验证模型的有效性，除了直观地与真图进行视觉对比外，还需要利用评价指标定量分析模型的性能。当前常用客观测量的方法有：

a) 峰值信噪比(peak signal to noise ratio, PSNR)[4]，是一种简单且广泛使用的 SISR 测量方法，它通过均方差(MSE)进行定义。假设两个 $m\times n$ 单色图像 I 和 K，如果一个为另外一个的噪声近似，那么 PSNR 定义为

$$PSNR = 10\bullet\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\bullet\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{10}$$

MSE 定义为

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left\|I(i,j) - K(i,j)\right\|^2 \tag{11}$$

其中：MAX_I 是表示图像点颜色的最大数值。两幅图像间的 PSNR 值(单位 dB)越高，则重建图像相对于高分辨率图像失真得越少。

b) 结构相似性 (structural similarity index method, SSIM)[69],采用更加直接的方法来比较重建图像和参考图像的结构。SSIM 测量测量由三种对比模块组成，分别为亮度、对比度、结构。假设给定两幅大小为 MN 的图像 X、Y，其中 X 和 Y 的均值、方差及 X 和 Y 的协方差分别用 u_x、u_y、δ_x^2、δ_y^2、δ_{xy} 表示。定义亮度、对比度、结构的比较函数分别为

$$l(X,Y) = \frac{2u_xu_y + c1}{u_x^2 + u_y^2 + c1} \tag{12}$$

$$c(X,Y) = \frac{2\delta_x\delta_y + c2}{\delta_x^2 + \delta_y^2 + c2} \tag{13}$$

$$s(X,Y) = \frac{\delta_{xy} + c3}{u_xu_y + c3} \tag{14}$$

这三个成分因素综合起来就是 SSIM 指标如式(15)所示。

$$SSIM(X,Y) = \left[l(X,Y)\right]^{\alpha}\left[c(X,Y)\right]^{\beta}\left[s(X,Y)\right]^{\gamma} \tag{15}$$

SSIM 值越趋近于 1，则两幅图相似度越高。

c) 信息保真标准(information fidelity criterion, IFC)，通过计算待评图像与参考图之间的信息来衡量待评图像的质量优劣，它的判断方法与 PSNR 相同。

3.2 实验结果对比

本节针对多种单图像超分辨率重建的深度学习模型在不同数据集下进行实验对比和数据分析。实验所用数据集包括 Set5、Set14、image91 和 DIV2K 数据集。实验所有环境如表

1 所示。

表 1 实验环境的配置参数

Table 1 Experimental environment configuration parameters

参数	数值	参数	数值
操作系统	Ubuntu 14.04	CUDA 版本	CUDA8.0
CPU	Intel Xeon 2.10GHz	cuDNN 版本	cudnn-7.5
GPU	GeForceGTX1080/8G	tensorflow-GPU	1.2.0
RAM	32G/DDR3/2.10GHz	keras	2.0.0

实验对比了七种 SISR 算法，包括传统的 SISR 算法(Bicubic)、基于插值预处理的模型(如 SRCNN 和 DRCN)、基于原始图像的模型(ESPCN、FSRCNN)、基于分层特征学习的模型(SRDenseNet)以及基于高频细节信息的模型(SRGAN)。

图 6 可视化地展示了部分 Set5 数据集在五种算法上的运行结果。从左到右依次为 BICUBIC、SRCNN、ESPCN、DRCN、SRGAN。第一列为原始图像，可以发现 BICUBIC 方法，只是简单进行了插值方法，较原始图像可视化效果较差，SRCNN 算法在插值放大的基础上进行了卷积操作，所以其可视化效果较 BICUBIC 算法有所提高，但是因为只有三个卷积层，所以输出结果存在边缘过于平滑问题。可视化结果较好的为 SRGAN 算法，其次是 DRCN，说明随着网络层数的加深，提取的信息越来越多，图像重建效果也越来越好。

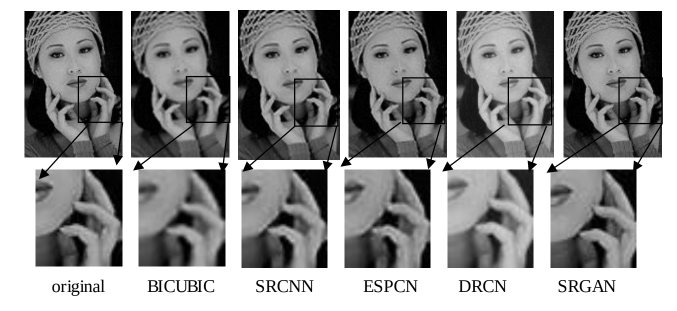


图 6 部分方法在 Set5 数据集上的可视化结果

Fig.6 Visualization results of partical methods on Set5 dataset

表 2 给出了相同的 Set5 数据集上，原图放大四倍情况下不同算法的定量分析结果。从表中可以看出，在所有算法较非深度学习算法(Bicubic)其 PSNR、SSIM 值和 MOS 值都有所提高，这也证实了深度学习算法在 SISR 问题上的优势。

就时间方面，传统的方法因其结构较为简单，用时最短，但是其图像失真较为严重。现有深度学习方法中，ESPCN 和 FSRCNN 是在 SRCNN 的基础上对结构进行的改进，其速度较 SRCNN 分别提高了 0.089 s 和 0.034 s，DRCN 和 SRDenseNet，SRGAN 因为其网络层数较多，所以耗时较久。

就 PSNR 和 SSIM 而言，所有深度学习的方法，因为其端到端学习的优势，所以 PSNR 和 SSIM 值比传统方法均有所提高。其中 SRDenseNet 算法因为结合了密集残差网络的思想，充分利用了层与层之间的特征信息，所以其 PSNR 和 SSIM 的值在七组实验中均属于最高，分别为 32.02 dB 和 0.893 4。

在视觉方面，MOS 值最高的是 SRGAN 方法，其值高达 3.58，这是因为 SRGAN 方法充分考虑的原始图像的高频细节信息，通过利用内容损失函数在视觉上生成更符合真实图像的超分辨率图像，但是其 PSNR 和 SSIM 值有待提高。总之，现有算法依然没有达到理想的效果。

综上所述，上述算法综合性能较优的是 SRDenseNet 算法，但是其时间和视觉效果仍有待提高。相对于传统的单图像超分辨率重建模型，通过利用循环神经网络和残差网络增加了模型的深度，提高模型的性能；通过密集网络和生成式对抗网络充分利用图像的层间特征和高频特征，达到更好的视觉效果，间接说明了研究深度学习算法的必要性。在未来还需进一步深入学习深度学习的算法，探索更好的网络模型，进而达到更好的单图像超分辨率重建效果，这也是单图像超分辨率重建的一个研究重点和难点。

表 2 不同算法的定量结果比较

Table 2 Comparison of quantitative results of different algorithms

Set5	Bicubic	SRCNN	DRCN	ESPCN	FSRCNN	SRDenseNet	SRGAN
PSNR	28.42dB	30.09dB	31.52dB	30.90dB	30.71dB	**32.02dB**	29.90dB
SSIM	0.8211	0.8627	0.8938	0.8784	0.8657	**0.8934**	0.8472
MOS	1.97	2.57	3.26	2.89	2.74	3.37	**3.58**
Time	0.016s	0.136s	0.298s	0.047s	0.102s	0.3194s	0.3264s

4 小结

单图像超分辨率重建是计算机视觉领域的一项基础研究工作，可以作为后续许多研究的先导性操作，具有十分重要的理论研究意义和实际应用价值。本章将进一步总结梳理基于深度学习的单图像超分辨率重建的关键问题，并对未来的发展趋势进行展望。

a) 现有算法优化。现有的单图像超分辨率算法虽然在 PSNR 和 SSIM 值上有了很大的提升，但是其可视化效果并没有达到人们的预期，具有较大的发展空间。GAN 网络虽然在一定程度上提高了可视化的效果，但其 PSNR 和 SSIM 值较低，且 GAN 本身存在不稳定的问题。所以未来的工作应考虑在已改进 GAN 网络的基础上结合当前较新的算法如 ResneXt 算法、indRNN(独立循环神经网络)算法，提高训练结果可视化，提高训练结果定量分析值，保证网络模型的稳定性。

b)理论知识的指导。单图像超分辨率重建的理论研究与数字图像领域联系较为密切，需深入理解分析领域知识，将其理论思想如超分辨率重建的理论极限、图像高低分辨率信息之间的关系等借鉴到单图像超分辨率重建的研究中，会是未来的一个热点研究方向。

c)实际场景应用需求。人工智能火热发展的同时，智能交通、视频监控、光学文字识别等方面的应用需求日益增加，因此，如何结合实际应用场景，将特定领域的先验信息与深度网络结构、代价函数以及训练方式结合是一个有潜力的研究方向。

d)新的评价标准的制定。现有的主观评价方法不仅繁琐且昂贵，需要大量人力，而且也无法应用部署到基于输出质量反馈实时自我调整的自动系统中。基于客观评价的标准则大多依赖于图像的均方误差，利用了像素之间的相似性，没有充分考虑图像的内容损失，所以有必要综合考虑时间、均方误差、内容损失三者之间的关系，在现有评价标准的基础上制定一套更加准确、灵活的新衡量标准，这是未来极具价值的研究方向。

总之，在未来很长的一段时间内，基于深度学习在单图像超分辨率重建将会取得更大的进展，因此仍需研究者提出更多有创新性实用性的模型和方法。

5 结束语

本文系统总结了目前基于深度学习的图像超分辨率重建中的相关模型，结合对比实验分析了现有模型的存在的问题，并提出了一些具有实际参考价值的建议，总结了未来发展趋势和所面临的挑战，为深入研究奠定了基础。

参考文献：

[1] 刘村，李元祥，周拥军，等. 基于卷积神经网络的视频图像超分辨率重建方法 [J]. 计算机应用研究，2019，36(4): 1-8. (Liu Cun, Li Yuanxiang, Zhou Yongjun, *et al.* Video image super-resolution reconstruction method based on convolutional neural network [J]. Application Research of Computers, 2019，36 (4): 1-8.)

[2] Duchon C E. Lanczos filtering in one and two dimensions [J]. Journal of Applied Meteorology, 1979, 18 (8): 1016-1022.

[3] Irani M, Peleg S. Motion analysis for image enhancement: resolution, occlusion, and transparency [J]. Journal of Visual Communications & Image Representation, 1993, 4 (4): 324-335.)

[4] 苏衡，周杰，张志浩. 超分辨率图像重建方法综述 [J]. 自动化学报，2013, 39 (8): 1202-1213.) (Su Heng, Zhou Jie, Zhang Zhihao. Survey of super-resolution image reconstruction methods [J]. Acta Automatica Sinica, 2013, 39 (8): 1202-1213.)

[5] Bauschke H H, Borwein J M. On projection algorithms for solving convex feasibility problems [J]. Siam Review, 1996, 38 (3): 367-426.)

[6] Chang Hong, Yeung D Y, Xiong Yimin. Super-resolution through neighbor embedding [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2004.

[7] Yang Jianchao, Wright J, Thomas S, *et al.* Image super-resolution via sparse representation [J]. IEEE Trans on Image Processing, 2010, 19 (11): 2861-2873.

[8] Wang Zhaowen, Liu Ding, Yang Jianchao, *et al.* Deep networks for image super-resolution with sparse prior [C]//Proc of IEEE International Conference on Computer Vision. [S.l.]:IEEE Computer Society, 2015: 370-378.

[9] Cui Zhen, Chang Hong, Shan Shiguang, *et al.* Deep network cascade for image super-resolution [C]//Proc of European Conference on Computer Vision. Berlin: Springer, 2014: 49-64.

[10] Dong Chao, Loy C C, He Kaiming, *et al.* Learning a deep convolutional network for image super-resolution [C]// Proc of European Conference on Computer Vision. Berlin: Springer, 2014: 184-199.

[11] Osendorfer C, Soyer H, Smagt P V D. Image super-resolution with fast approximate convolutional sparse coding [C]// Proc of International Conference on Neural Information Processing. Berlin: Springer, 2014: 250-257.

[12] Yang Jianchao, Wang Zhaowen, Lin Zhe, *et al.* Coupled dictionary training for image super-resolution [J]. IEEE Trans on Image Processing. 2012, 21 (8): 3467-3478.

[13] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18 (7): 1527-1554.

[14] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521 (7553): 436.

[15] Zouxy09. Deep learning（深度学习）学习笔记整理系列之（八）. [EB/OL] (2018-08-19). [2018-09-11]. https://blog.csdn.net/zouxy09/article/details/10077055. (Zouxy09.Deep learning (deep learning) study note finishing series (eight). [EB/OL]. (2018-08-19). [2018-09-11]. https://blog.csdn.net/zouxy09/article/details/10077055.)

[16] 周飞燕，金林鹏，董军. 卷积神经网络研究综述 [J]. 计算机学报，2017, 40 (6): 1229-1251. (Zhou Feiyan, Jin Linpeng, Dong Jun. Review of convolutional neural network [J]. Chinese Journal of Computers. 2017, 40 (6): 1229-1251.)

[17] Lecun Y, Boser B, Denker J S, *et al.* Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1 (4): 541-551.

[18] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]//Proc of International Conference on Neural Information Processing Systems. [S.l.]:Curran Associates Inc, 2012: 1097-1105.

[19] Bakhoda A, Yuan G L, Fung W W L, *et al.* Analyzing CUDA workloads using a detailed GPU simulator[C]//Proc of IEEE International Symposium on Performance Analysis of Systems and Software. Piscataway, NJ: IEEE Press, 2009: 163-174.

[20] Hinton G E. Rectified linear units improve restricted boltzmann machines vinod nair [C]//Proc of International Conference on International Conference on Machine Learning. 2010.

[21] Deng Jia, Dong Wei, Socher R, *et al.* ImageNet: a large-scale hierarchical image database [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2009: 248-255.

[22] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]//Proc of IEEE International Conference on Computer Vision. [S.l.]:IEEE Computer Society, 2016: 770-778.

[23] 石仕伟. 基于深度学习的视频行为识别研究 [D]. 杭州：浙江大学，2018. (Shi Shiwei. Research on video behavior recognition based on deep learning [D]. Hangzhou:Zhejiang University, 2018) .

[24] Targ S, Almeida D, Lyman K. Resnet in resnet: generalizing residual architectures [J]. ResearchGate. 2016 (3): 140.

[25] 郝旭政，柴争义. 一种改进的深度残差网络行人检测方法 [J]. 计算机应用研究，2019,36(6): 1-3. (Hao Xuzheng, Chai Zhengyi. Improved pedestrian detection method based on depth residual networ [J]. Application Research of Computer, 2019,36 (6): 1-3.)

[26] 杨祎玥，伏潜，万定生. 基于深度循环神经网络的时间序列预测模型 [J]. 计算机技术与发展. 2017, 27 (3): 35-38. (Yang yiwei, Fu qian, Wan Dingsheng. A prediction model for time series based on deep recurrent neural network [J]. Computer Technology and Development, 2017, 27 (3): 35-38.)

[27] Auli M, Galley M, Quirk C, *et al.* Joint language and translation modeling with recurrent neural networks [J]. American Journal of Psychoanalysis, 2013, 74 (2): 212-3.

[28] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks [C]//Proc of International Conference on Machine Learning. 2014: 1764-1772.

[29] Karpathy A, Li FeiFei. Deep visual-semantic alignments for generating image descriptions [C]//Proc of IEEE International Conference on Computer Vision. [S.l.]:IEEE Computer Society, 2015: 3128-3137.

[30] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets [C]//Proc of International Conference on Neural Information Processing Systems. [S.l.]:MIT Press, 2014: 2672-2680.

[31] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521 (7553): 436.

[32] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [C]// Proc of ICLR.2016.

[33] Chen Xi, Duan Yan, Houthooft Rein, *et al.* InfoGAN: interpretable representation learning by information maximizing generative adversarial nets [C]// Proc of Neural Information Processing Systems. 2016.

[34] Zhao Junbo, Mathieu M, LeCun Y. Energy-based generative adversarial

network [J]. ResearchGate. 2016 (6): 273.

[35] Salimans T, Goodfellow I, Zaremba W, *et al.* Improved techniques for training GANs [J]. ResearchGate. 2016 (6): 1196.

[36] Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks [J]. ResearchGate. 2017 (1): 279.

[37] Gou Chao, Yue Wu, Wang Kang, *et al.* Learning-by-synthesis for accurate eye detection [C]// Proc of IEEE International Conference on Pattern Recognition. 2017: 3362-3367.

[38] Santana E, Hotz G. Learning a driving simulator [J]. ResearchGate. 2016 (8): 145.

[39] Li Jiwei, Monroe W, Shi Tianlin, *et al.* Adversarial learning for neural dialogue generation [J]. ResearchGate. 2017 (1): 152.

[40] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks [C]// Proc of the 28th International Conference on Neural Information Processing Systems. 2015: 2377-2385.

[41] Gao Huang, Yu Sun, Zhuang Liu, *et al.* Deep networks with stochastic depth [C]//Proc of European Conference on Computer Vision. Cham:Springer, 2016.

[42] Larsson G, Maire M, Shakhnarovich G. FractalNet: ultra-deep neural networks without residuals [J]. ResearchGate. 2016 (5): 298.

[43] Huang Gao, Liu Zhuang, Van der Maaten L, *et al.* Densely connected convolutional networks [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]:IEEE Computer Society, 2017: 2261-2269.

[44] Krešo I, Oršić M, Bevandić P, *et al.* Robust semantic segmentation with ladder-densenet models [J]. ResearchGate, 2018 (6): 36.

[45] Park S, Jeong Y, Kim H S. Multi-resolution DenseNet based acoustic models for reverberant speech recognition [J]. Phonetics & Speech Sciences, 2018, 10 (1): 33-38.

[46] Liu Wenqi, Zeng Kun. SparseNet: a sparse densenet for image classification [J]. ResearchGate. 2018 (4): 55.

[47] Dong Chao, Loy C C, He Kaiming, *et al.* Learning a deep convolutional network for image super-Resolution [M]//Computer Vision–ECCV 2014. [S.l.]:Springer International Publishing, 2014: 184-199.

[48] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks [Z]. 2015: 1646-1654.

[49] Lai Weisheng, Huang Jiabin, Ahuja N, *et al.* Deep laplacian pyramid networks for fast and accurate super-resolution [C]//Proc of IEEE International Conference on Computer Vision. [S.l.]:IEEE Computer Society, 2017: 5835-5843.

[50] Tai Ying, Yang Jian, Liu Xiaoming. Image super-resolution via deep recursive residual network [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]:IEEE Computer Society, 2017: 2790-2798.

[51] Zhang Kai, Zuo Wangmeng, Gu Shuhang, *et al.* Learning deep CNN denoiser prior for image restoration [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2808-2817.

[52] Tai Ying, Yang Jian, Liu Xiaoming, *et al.* MemNet: a persistent memory network for image restoration [C]//Proc of IEEE International Conference on Computer Vision. [S.l.]:IEEE Computer Society, 2017: 4549-4557.

[53] Li Sumei, Fan Ru, Lei Guoqing, *et al.* A two-channel convolutional neural network for image super-resolution [J]. Neurocomputing, 2018 (275): 267-277.

[54] Dong Chao, Loy C C , Tang Xiaoou. Accelerating the super-resolution convolutional neural network [C]//Proc of European Conference on Computer Vision.Berlin: Springer, 2016: 391-407..

[55] Mao Xiaojiao, Shen Chunhua, Yang Yubin. Image restoration using convolutional auto-encoders with symmetric skip connections [J]. ResearchGate,2016 (6): 391.

[56] Lai Weisheng, Huang Jiabin, Ahuja N, *et al.* Deep laplacian pyramid networks for fast and accurate super-resolution [C]//Proc of IEEE International Conference on Computer Vision. IEEE Computer Society, 2017: 5835-5843.

[57] Shi Wenzhe, Caballero Jose, Huszár Ferenc, *et al.* Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]//Proc of IEEE International Conference on Computer Vision. [S.l.]: IEEE Computer Society, 2016.

[58] Ledig C, Theis L, Huszar F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network [J]. ResearchGate. 2016 (9): 105-114.

[59] Lim B, Son S, Kim H, *et al.* Enhanced deep residual networks for single image super-resolution [C]//Proc of Computer Vision and Pattern Recognition Workshops. 2017: 1132-1140.

[60] Tong Tong, Li Gen, Liu Xiejie, *et al.* Image super-resolution using dense skip connections [C]// Proc of IEEE International Conference on Computer Vision. [S.l.]: IEEE Computer Society, 2017: 4809-4817.

[61] Zhang Yulun, Tian Yapeng, Kong Yu, *et al.* Residual dense network for image super-resolution [J]. ResearchGate. 2018 (2): 180.

[62] Zhou Fuqiang, Li Xiaojie, Li Zuoxin. High-frequency details enhancing densenet for super-resolution [J]. Neurocomputing, 2018 (290): 34-42.

[63] Ledig C, Theis L, Huszar F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network [J]. ResearchGate. 2016: 105-114.

[64] Bosch M, Gifford C M, Rodriguez P A. Super-resolution for overhead imagery using densenets and adversarial learning [C]// Proc of IEEE Winter Conference on Applications of Computer Vision. [S.l.]:IEEE Computer Society, 2018: 1414-1422.

[65] Zhang Dongyang, Shao Jie, Hu Gang, *et al.* Sharp and real image super-resolution using generative adversarial network [C]// Proc of International Conference on Neural Information Processing. 2017: 217-226.

[66] 杨文瀚，刘家瑛，夏思烽，等. 数据外补偿的深度网络超分辨率重建 [J]. 软件学报, 2018, 29 (4): 900-913. (Yang Wenhan, Liu Jiaying, Xia Sifeng, *et al.* Data-driven external compensation guided [J]. Journal of Software, 2018, 29 (4): 900-913.)

[67] Ostaszewska A, S. ŻebrowskaŁucyk. The method of increasing the accuracy of mean opinion score estimation in subjective quality evaluation [M]//Wearable and Autonomous Biomedical Devices and Systems for Smart Environment. Berlin: Springer, 2010.

[68] Zhou Wang, Bovik A C, Sheikh H R, *et al.* Image quality assessment: from error visibility to structural similarity[J]. IEEE Trans Image Process, 2004, 13 (4): 600-612.

[69] Sheikh H R, Bovik A C, Veciana G D. An information fidelity criterion for image quality assessment using natural scene statistics [J]. IEEE Trans Image Process, 2005, 14 (12): 2117-2128.